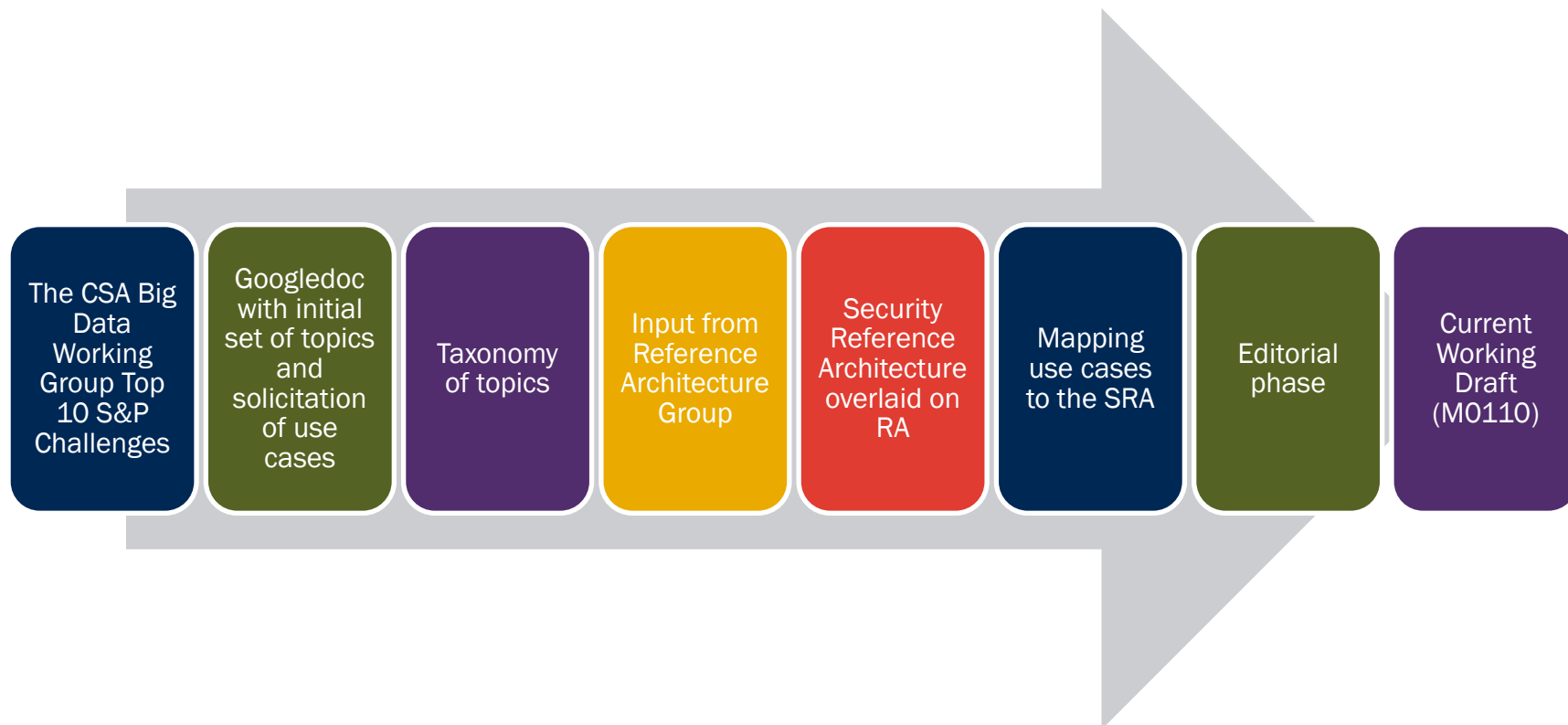


NIST Big Data Public Working Group

**Security and Privacy Requirements
March 18, 2014**

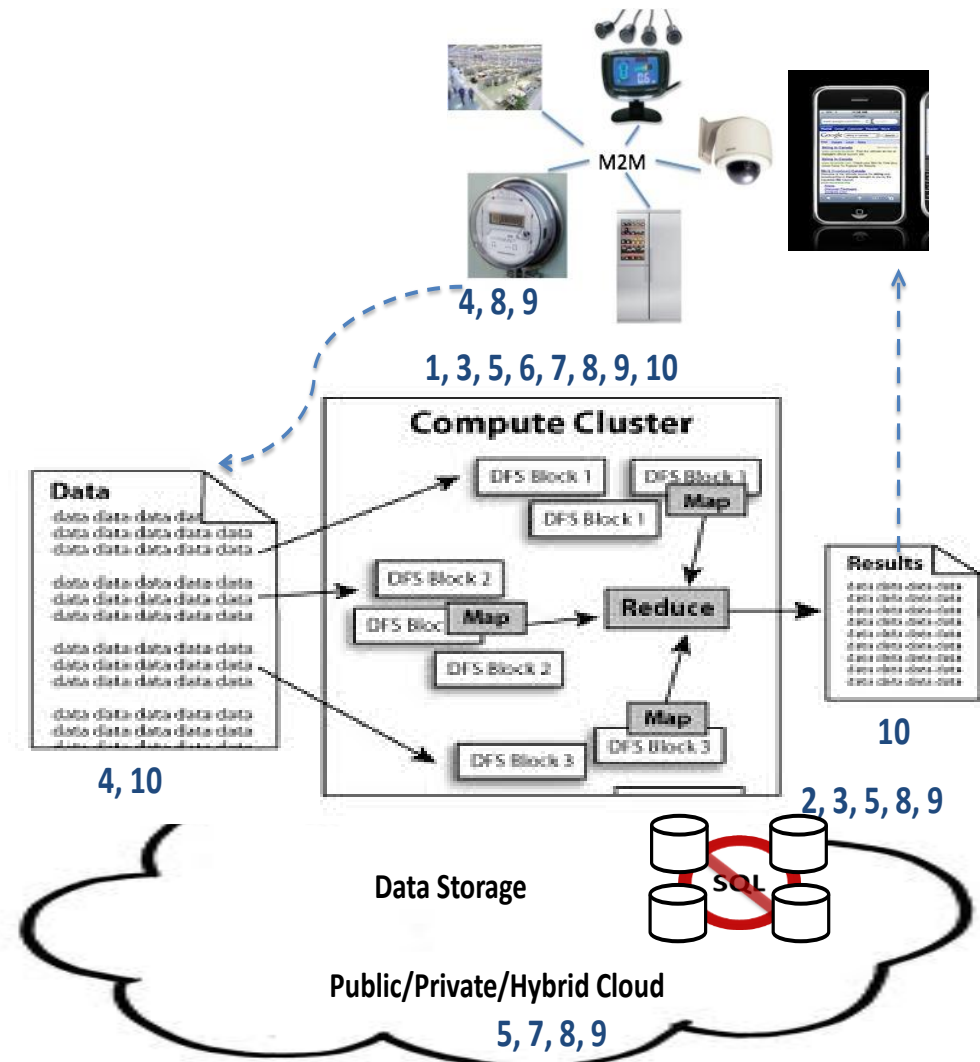
**Arnab Roy, Fujitsu
On behalf of the NIST BDWG S&P Subgroup**

Process

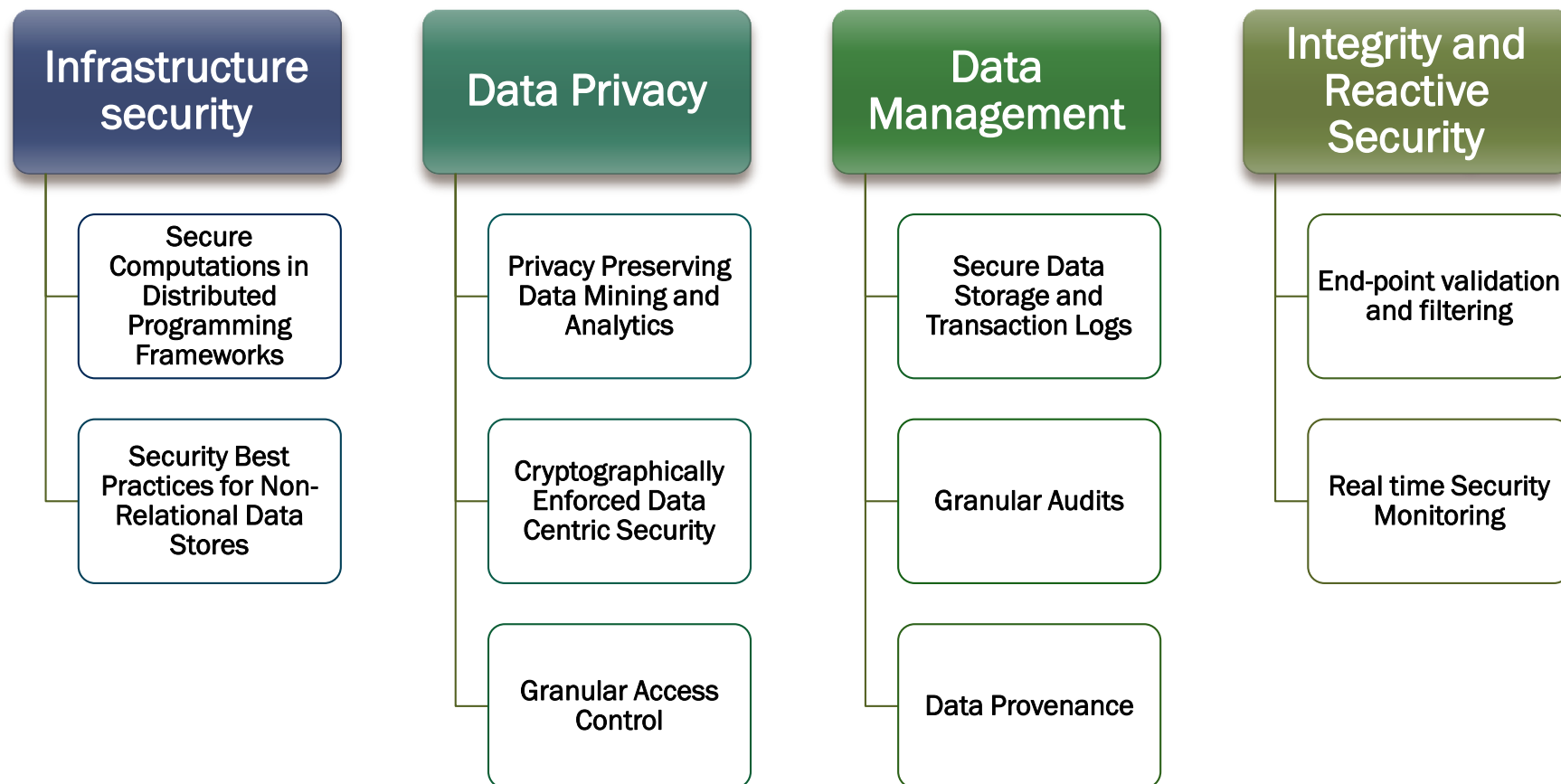


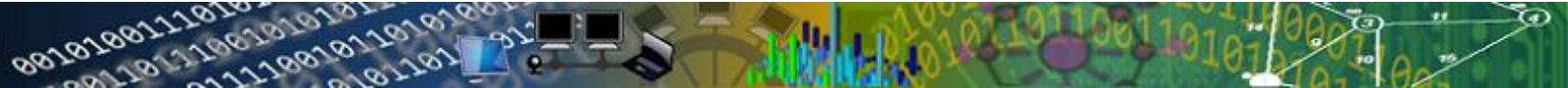
CSA BDWG: Top Ten Big Data Security and Privacy Challenges

- 1) Secure computations in distributed programming frameworks
- 2) Security best practices for non-relational datastores
- 3) Secure data storage and transactions logs
- 4) End-point input validation/filtering
- 5) Real time security monitoring
- 6) Scalable and composable privacy-preserving data mining and analytics
- 7) Cryptographically enforced access control and secure communication
- 8) Granular access control
- 9) Granular audits
- 10) Data provenance



Top 10 S&P Challenges: Classification





Taxonomy – conceptual axis

Privacy

Communication Privacy		
Data Confidentiality	Access Policies	Systems Crypto Enforced
Computing on Encrypted Data	Searching and Reporting Fully Homomorphic Encryption	
Secure Data Aggregation		
Key Management		

Provenance

End-point Input Validation	Syntactic Validation Semantic Validation	
Communication Integrity		
Authenticated Computations on Data	Trusted Platforms Crypto Enforced	
Granular Audits		
Control of Valuable Assets	Lifecycle Management Retention, Disposition, Hold Digital Rights Management	

System Health

Security against DoS	Construction of cryptographic protocols proactively resistant to DoS	
Big Data for Security	Analytics for Security Intelligence Data-driven Abuse Detection Event Detection Forensics	

Taxonomy – operational axis

Big Data Security and Privacy

Registration, Security Model and Policy Enforcement

- Device, User, Asset, Services, Applications registration
- Security Metadata Model
- Policy Enforcement

Identity and Access Management

- Virtualization Layer Identity
- Application Layer Identity
- End User Layer Identity Management
- Identity Provider
- Additional XACML Concepts

Data Governance

- Encryption and Key Management (including Multi-Key)
- Isolation/Containerization
- Storage Security
- Data Loss Prevention, Detection
- Web Services Gateway
- Data Transformation
- Data Lifecycle Management
- End Point Input Validation
- Digital Rights Management

Visibility and Infrastructure Management

- Threat and Vulnerability Management
- Monitoring, Alerting
- Mitigation
- Configuration Management
- Logging
- Malware Surveillance and Remediation
- Network Boundary Control
- Resiliency, Redundancy and Recovery

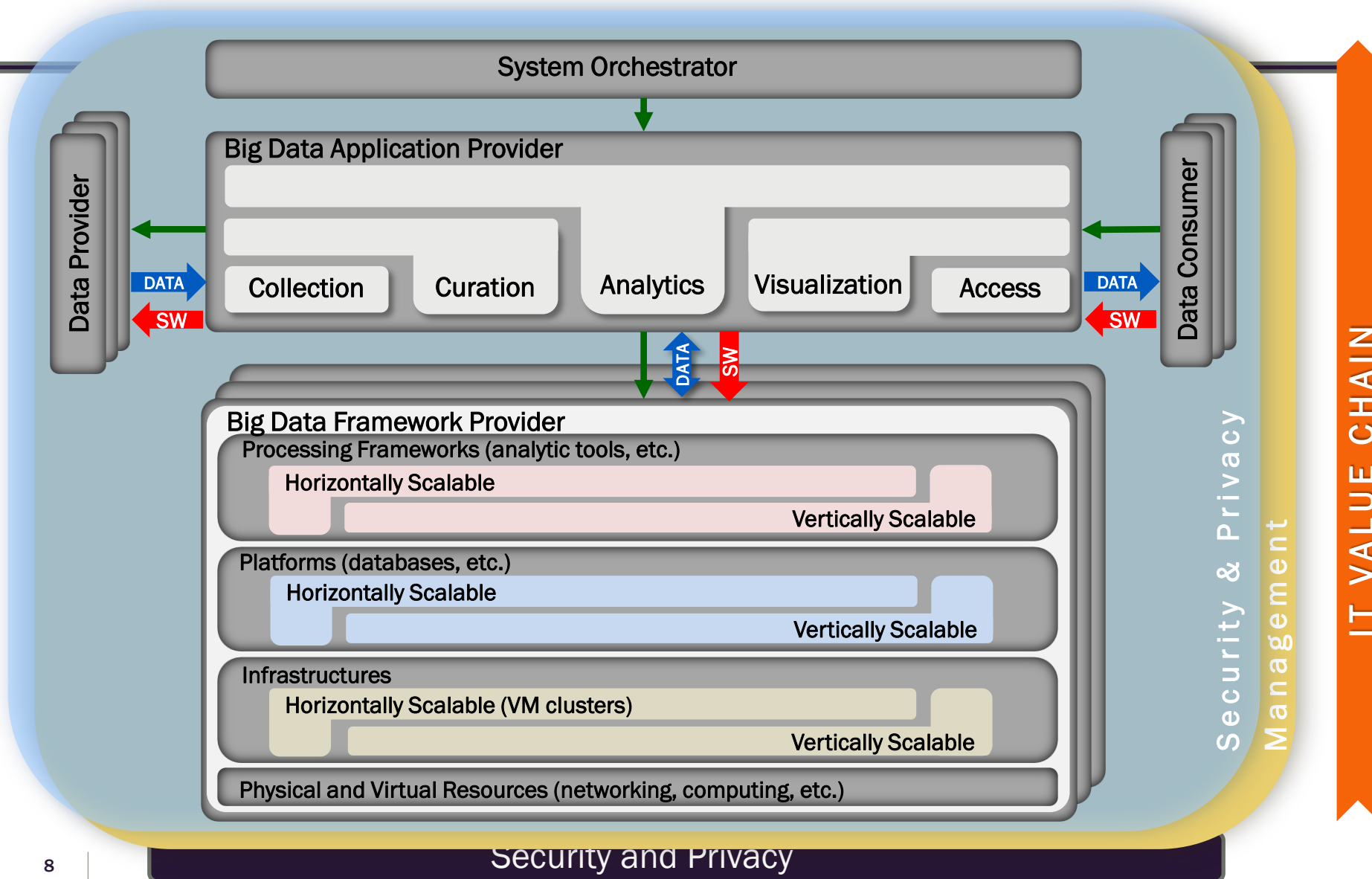
Risk and Accountability

- Accountability
- Compliance
- Forensics
- Business Risk Model

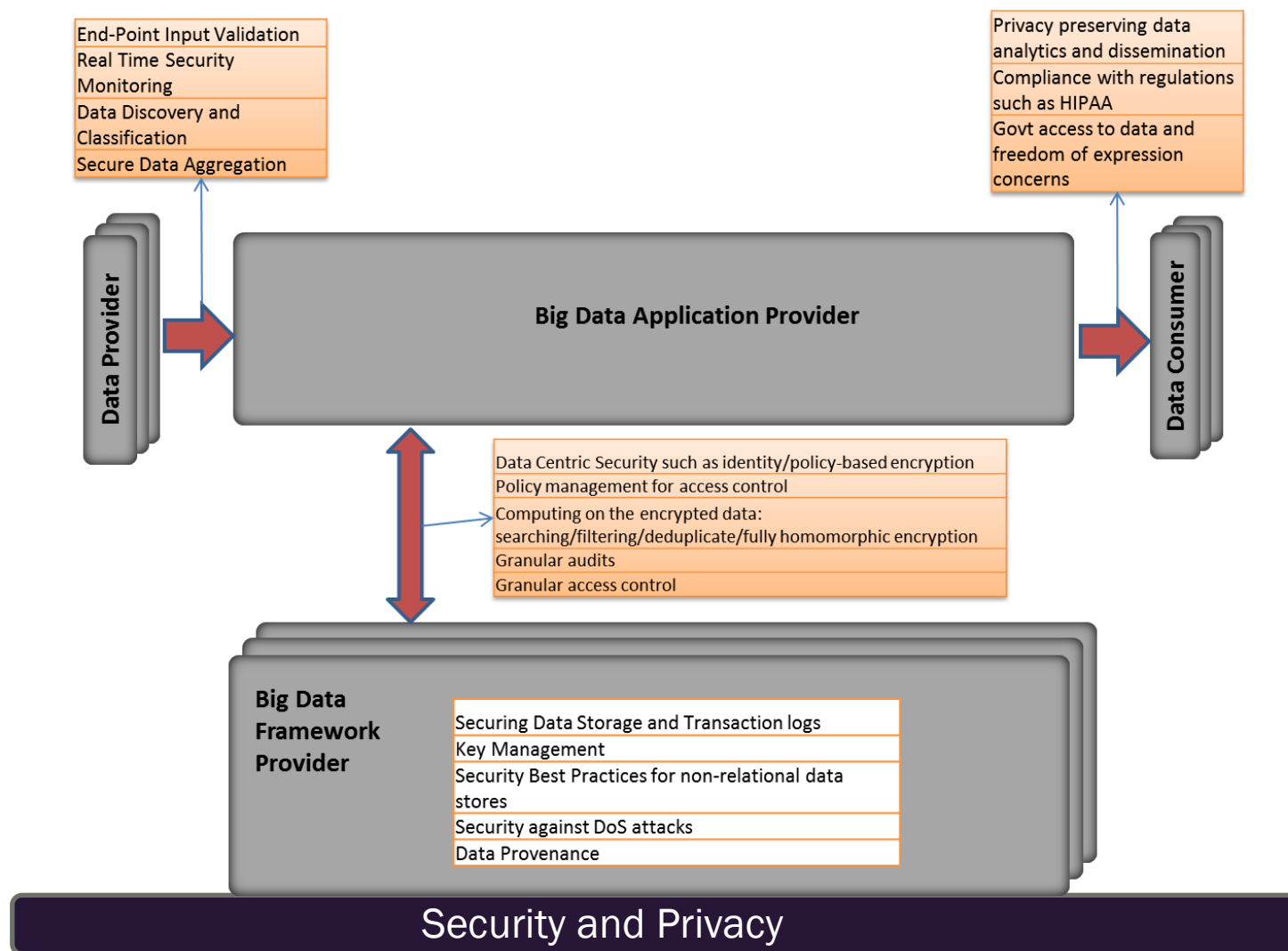
Use Cases

- **Retail/Marketing**
 - Modern Day Consumerism
 - Nielsen Homescan
 - Web Traffic Analysis
- **Healthcare**
 - Health Information Exchange
 - Genetic Privacy
 - Pharma Clinical Trial Data Sharing
- **Cyber-security**
- **Government**
 - Military
 - Education

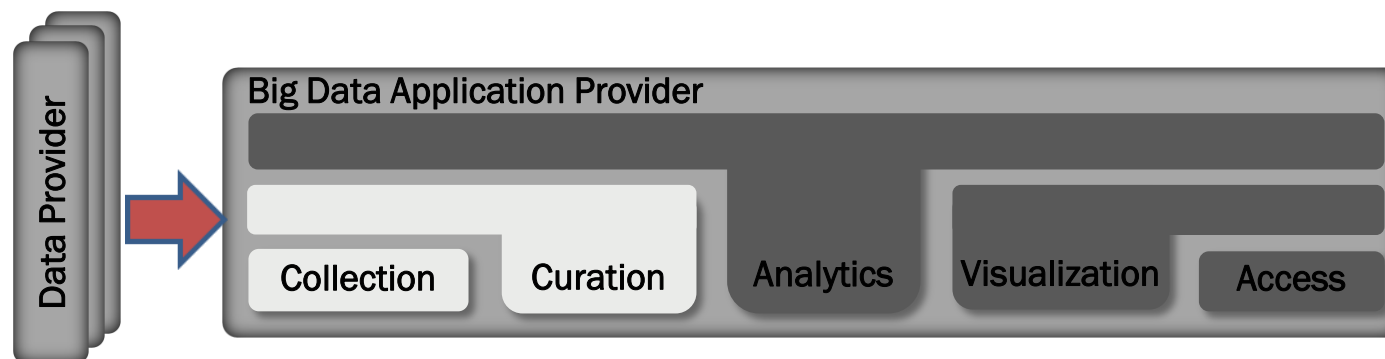
INFORMATION VALUE CHAIN



Big Data Security Reference Architecture

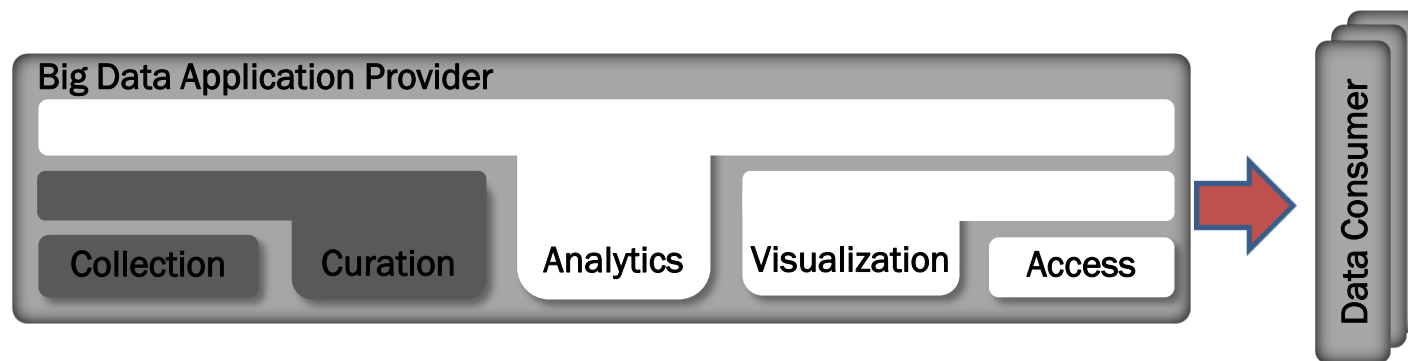


Interface of Data Providers -> BD App Provider



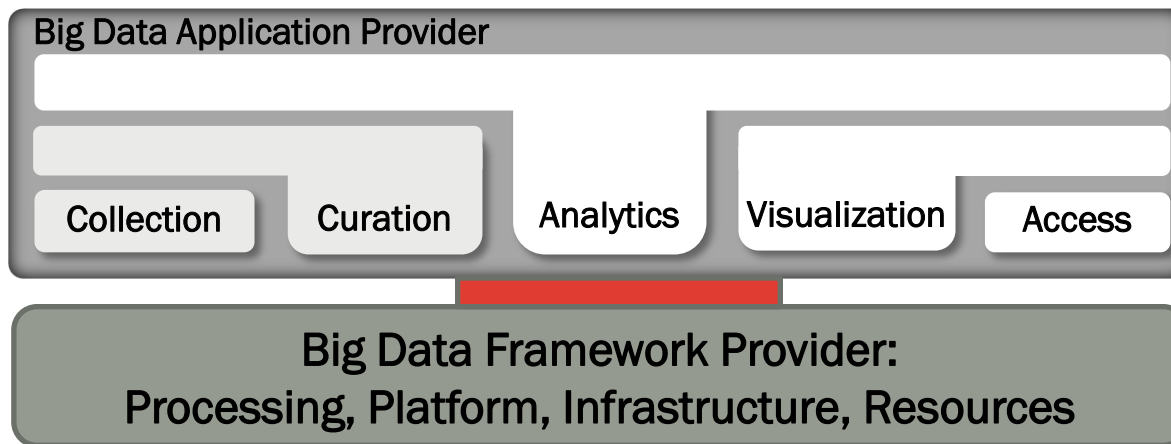
S&P Consideration	Health Info Exchange	Military UAV
End-Point Input Validation	Strong authentication, perhaps through X.509v3 certificates, potential leverage of SAFE bridge in lieu of general PKI	Need to secure sensor to prevent spoofing/stolen sensor streams
Real Time Security Monitoring	Validation of incoming records. May need to check for evidence of Informed Consent.	On-board & control station secondary sensor security monitoring
Data Discovery and Classification	Leverage HL7 and other standard formats opportunistically, but avoid attempts at schema normalization.	Varies from media-specific encoding to sophisticated situation-awareness enhancing fusion schemes.
Secure Data Aggregation	Clear text columns can be deduplicated, perhaps columns with deduplication.	Fusion challenges range from simple to complex.

Interface of BD App Provider -> Data Consumer



S&P Consideration	Health Info Exchange	Military UAV
Privacy preserving data analytics and dissemination	Searching on encrypted data. Determine if drug administered will generate an adverse reaction, without breaking the double blind.	Geospatial constraints: cannot surveil beyond a UTM. Military secrecy: target, point of origin privacy.
Compliance with regulations	HIPAA security and privacy will require detailed accounting of access to HER data.	Numerous. Also standards issues.
Govt access to data and freedom of expression concerns	CDC, Law Enforcement, Subpoenas and Warrants. Access may be toggled based on occurrence of a pandemic or receipt of a warrant.	Google lawsuit over streetview.

Interface of BD App Provider -> BD Framework Provider



S&P Consideration	Health Info Exchange	Military UAV
Policy based encryption	Row-level and Column-level Encryption	Policy-based encryption, often dictated by legacy channel capacity/type.
Policy management for access control	Role-based and claim-based	Transformations tend to be made within DoD-contractor devised system schemes.
Computing on encrypted data	Privacy preserving access to relevant events, anomalies and trends.	Sometimes performed within vendor-supplied architectures, or by image-processing parallel architectures.
Audits	Facilitate HIPAA readiness, and HHS audits	CSO, IG audit.

Internal to BD Framework Provider

Big Data Framework Provider: Processing, Platform, Infrastructure, Resources

S&P Consideration	Health Info Exchange	Military UAV
Securing Data Stores and Transaction Logs	Need to be protected for integrity and for privacy, but also for establishing completeness, with an emphasis on availability.	The usual, plus data center security levels are tightly managed (e.g., field vs. battalion vs. HQ).
Security Best Practices for non-relational data	End-to-end encryption.	Not handled differently at present; this is changing in DoD.
Security against DoS attacks	Mandatory – availability is a compliance requirement.	DoD anti-jamming e-measures.
Data Provenance	Completeness and integrity of data with records of all accesses and modifications	Must track to sensor point in time configuration, metadata.



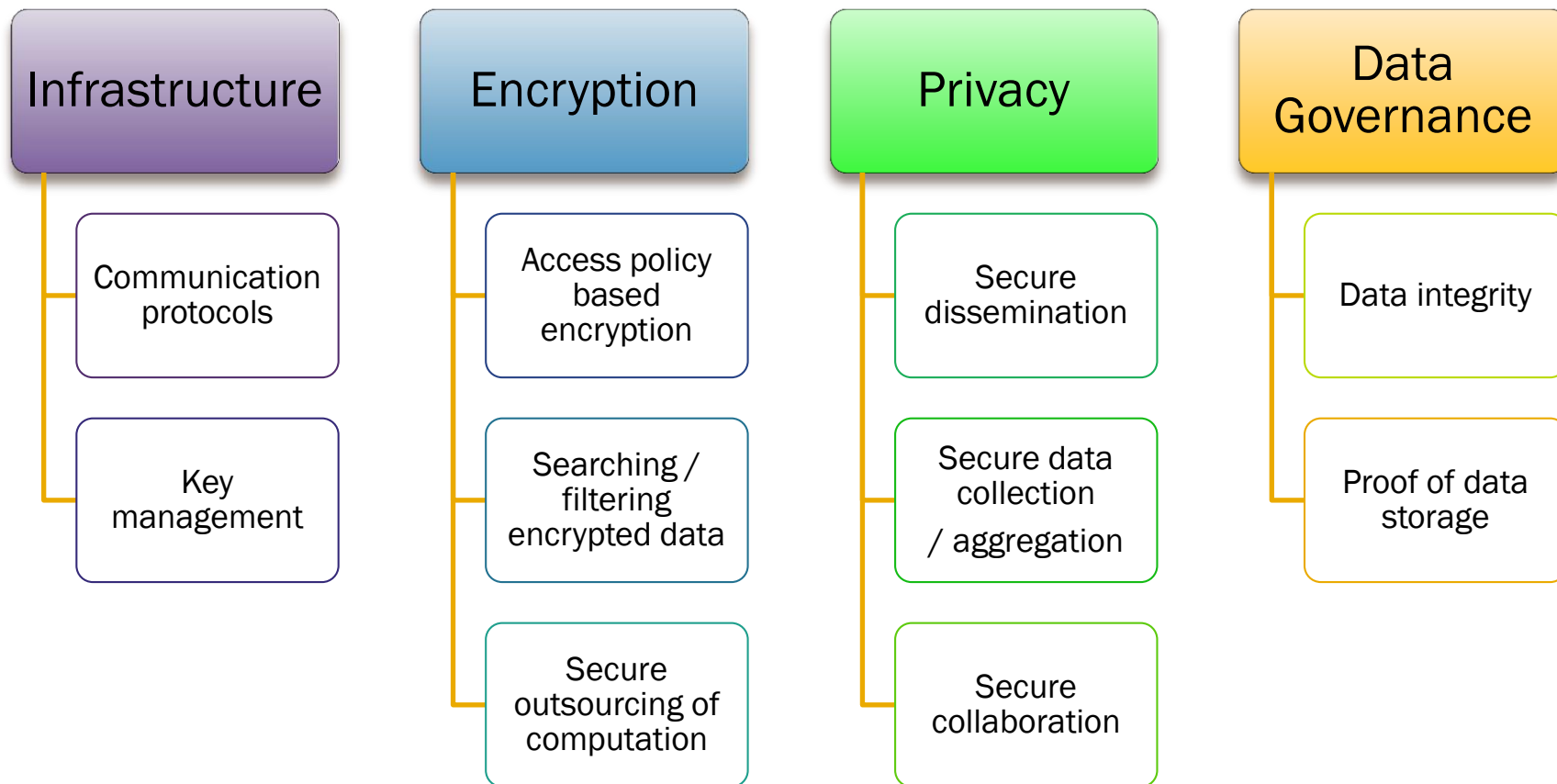
Next Steps

- Taxonomy to Reference Architecture Mapping
- Scope for Standards

Next steps: Cryptography and Privacy Enhancing Technologies

- **BIG**
 - Scale up existing solutions for volume, variety and velocity
 - Retarget to Big Data infrastructural shift
- **DATA**
 - Balance privacy and utility
 - Enable analytics and governance on encrypted data
 - Reconcile authentication and anonymity

Top 10 Challenges in Crypto and PET identified by CSA BDWG





Thank you!
